基于分布式压缩感知和散列函数的数据融合隐私保护算法 *

寇　兰，刘　宁，黄宏程†，张　艳

(重庆邮电大学 通信与信息工程学院, 重庆 400065)

摘　要：针对群智感知网络数据融合传输过程中隐私泄露、信息不完整、数据窜改等安全问题，提出了一种基于分布式压缩感知和散列函数的数据融合隐私保护算法。首先，采用分布式压缩感知方法对感知数据进行稀疏观测，去除冗余数据；其次，利用单向散列函数求取感知数据观测值的散列值，将其和不受限的伪装数据一起填充到感知数据观测值中，达到隐藏真实感知数据的目的；最后，在汇聚节点提取伪装数据之后，再次获取感知数据的散列值并验证数据的完整性。仿真结果表明，该算法兼顾了数据的机密性和完整性保护，同时大大降低了通信开销，在实际应用中具有很强的适用性和可扩展性。

关键词：隐私保护；分布式压缩感知；单向散列函数；群智感知网络；数据融合

中图分类号：TP309.2　　doi: 10.19734/j.issn.1001-3695.2018.06.0474

Data aggregation privacy protection algorithm based on distributed compressive sensing and hash function

Kou Lan, Liu Ning, Huang Hongcheng, Zhang Yan

(*School of Communication & Information Engineering*, *Chongqing University of Posts & Telecommunications*, *Chongqing* 400065, *China*)

Abstract: Aiming at the security problems existing in the process of the data aggregation and transmission in crowd sensing networks, such as privacy leakage, incomplete information, data tampering. this paper proposed a data aggregation privacy protection algorithm based on distributed compressive sensing and hash function. Firstly, it used distributed compressive sensing method to sparsely observe the sensed data and remove the redundant data. Then, it utilized one-way hash function to obtain hash value of the observation data and filled the hash value with the unconstrained camouflage data into the observation data of sensory data to reach the aim of concealing the true sensor data. Finally, after extracting the camouflage data at the sink node, it obtained the hash value of the observation data again to verify the integrity of data. Simulation results show that the algorithm takes into account the privacy preserving and integrity protecting of data, and also can reduce the communication overhead greatly, which means the strong applicability and scalability in practical applications.

Key words: privacy protection; distributed compressed sensing; one-way hash function; crowd sensing networks; data aggregation

0　引言

群智感知是以普通用户的移动设备（如智能手机、智能手表、平板电脑等）做为基本感知单元，通过系统互联网进行有意识和无意识的协作，进行感知任务的分发、感知数据的收集，完成复杂和大规模的社会感知任务[1]。相比于传统的传感器网络需要针对某一感知任务专门部署大量的传感器节点，群智感知网络可以直接利用用户随身携带的移动终端和现有的通信链路来构建，低成本实现大规模和细粒度的感知任务。因此，群智感知网络成为检测交通、地震和健康等信息的新型手段[2]。

在一些热点区域（比如商场、办公区域等），感知节点分布的较为密集，每个感知节点收集的数据与其他节点不可避免的存在较强的时空相关性。如果将原始感知数据直接传输到汇聚节点，那么对于冗余感知数据的不必要传输将会极大的增加感知网络的资源消耗[3]。而且，人们往往只关心感知数据的有效信息，并不需要大量的原始感知数据。因此，对原始感知数据压缩融合后传输，不仅可以极大程度的降低感知数据传输过程中的资源消耗，而且可以精确地提取原始海量感知数据的特

收稿日期：2018-06-14；**修回日期：**2018-07-23　　**基金项目：**重庆市科委基础与前沿研究项目（cstc2014jcyjA40039）

作者简介：寇兰（1963-），女，四川渠县人，副教授，硕士，主要研究方向为 D2D 通信、社会感知与社会计算；刘宁（1995-），女，河南周口人，硕士研究生，主要研究方向为群智感知网络；黄宏程（1979-），男（通信作者），河南南阳人，副教授，博士，主要研究方向为社会感知与社会计算（huanghc@cqupt.edu.cn）；张艳（1992-），女，湖北十堰人，硕士研究生，主要研究方向为群智感知网络.

征，提高感知平台对感知数据的收集效率。但是，无线传输链路的不可靠性和融合节点的不可信性使得感知数据容易受到对手窜改、窃听、重播、或注入虚假数据等，严重威胁了感知数据的安全[4]。因此，研究群智感知网络中的隐私保护，尤其是在数据融合阶段对感知数据的机密性和完整性保护，具有重要的研究意义。

1　相关工作

至今已有不少文献对群智感知网络数据融合传输过程中的隐私保护进行了研究。文献[5]提出了一种轻量级的定向数据融合方案，基于距离生成网络拓扑结构，以平衡簇头的能耗，由节点的私有因子和原始数据组成复数结构，并通过加性同态加密方法加密，实现了数据融合且不需要任何解密，该方案既保证了数据融合时的隐私保护性能又降低了计算开销。文献[6]提出了一种基于信任的数据融合协议，通过对节点行为的观察，协议计算、监视和评估节点的信任值，及时检测和排除受损节点，该方案能够有效的降低节点能耗并提升数据传输的可靠性。文献[7]通过使用简单的技术共享数据的加密内容但是不向其他节点显示真实数据和密钥，基站识别融合节点簇内相关的不信任节点，并且仅对异常感知数据的中间节点进行数据的重传，该方案中由于中间节点不需要逐跳进行加密解密操作，计算开销减小，中间融合节点不知道感知数据的真实内容，可以有效抵御内部攻击。

文献[8]提出了基于分簇的隐私保护数据融合方法，根据汇聚节点的指令，感知节点将生成的受限伪装数据和不受限制的伪装数据共同填充到感知数据观测值的对应位置，该方法加强了对数据的隐私保护性能。文献[9]针对分布式智能电表数据提出了基于傅里叶扰动算法和小波扰动算法的数据融合隐私保护数据系统，利用指数 ELGamal 加密机制实现用户和汇聚节点之间的安全通信，其中的分布式差分隐私机制依据高斯原理生成分布式噪声，与传统的隐私保护系统相比，保证了每个用户的差异隐私，但是该方案对于其他类型的感知数据不具备适用性。

为了保护感知数据的隐私，同时以较低的通信开销对数据进行融合，文献[10]提出了基于分布式压缩感知的隐私保护数据融合算法（distributed compressed sensing-based privacy-preserving data aggregation mechanism, DCSPDA），利用分布式压缩感知对原始的感知数据进行压缩和观测，以较小的通信开销实现了感知数据的隐私保护，但是缺乏考虑数据的完整性验证。文献[11]引入正负对来混淆真实感知数据，并通过混淆因子确定传感器需要产生的正负对数量，在数据融合阶段，采用正负中和策略和时间片分配机制，降低了通信开销和碰撞率，但是仍缺乏对感知数据的完整性保护。

基于安全多方计算的协议使节点交换种子并进行联合计算，有效保护了数据的隐私，但是在种子交换过程中，计算量和通信量随着涉及的节点数量而增加[12]。文献[13]提出了一种基于安全多方计算的车载网隐私保护方案，通过结合线性方程组理论、匿名认证思想和茫然传输协议与传统的公钥密码算法相比，大幅度降低了算法的计算复杂度。文献[14]基于安全多方计算的思想，对分簇数据融合隐私保护算法进行改进，簇头选择协作节点配合完成隐私数据融合工作，大大降低了计算量和通信量。

为了兼顾感知数据的机密性和完整性保护，文献[15]提出了隐私数据融合的完整性保护算法（integrity-protecting private data aggregation, iPDA）通过数据切片和组装计算来实现数据融合过程中的隐私保护，构建两颗融合树，并通过其中一颗用来监督融合结果是否完整。但是该算法只对部分攻击有效，例如对于攻击者同时窜改两棵树融合结果时的攻击，该算法便无法检测出。文献[16]提出了一种可检测数据完整性的隐私数据融合算法（integrity- checking private data aggregation，ICKPDA），该算法通过将密钥填充在感知数据中，并将填充密钥后的数据进行切片对感知数据进行隐私保护，利用密钥之间的关联性对切片在汇聚节点的重组数据进行完整性验证。该算法一定程度上兼顾了数据机密性和完整性，但是产生了较大的通信开销。

感知数据在融合传输时，应兼顾感知数据的机密性和完整性保护，同时需要尽量降低通信开销以满足实际场景的应用。现有的隐私保护方法大都以数据的机密性保护为中心，忽略了感知数据的完整性，而能够兼顾数据机密性和完整性的隐私保护方案通信开销较大，为了解决这个问题，本文提出一种基于分布式压缩感知和散列函数的数据融合隐私保护算法 (data aggregation privacy-preserving algorithm based on distributed compressed sensing and hash function, DAP-DCSHF），采用分布式压缩感知对原始感知数据进行稀疏和观测，降低网络的通信开销。采用单向散列函数对稀疏观测值求散列值，并将散列值和根据隐私保护需求随机生成的不受限伪装数据一起填充在观测值的零值位置，增强感知数据的机密性保护。将散列值和不受限伪装数据的填充位置信息单独加密发送给汇聚节点，最后汇聚节点将收到的隐私保护数据集中的伪装数据剔除后，再次利用散列函数求取散列值，通过比较两次散列值是否一致，验证感知数据的完整性。

2　基于分布式压缩感知和散列函数的数据融合隐私保护算法

本文提出基于分布式压缩感知和散列函数的数据融合隐私保护算法，在基于分布式压缩感知的隐私保护数据融合算法（DCSPDA）的基础上进行改进，利用单向散列函数对稀疏观测值求散列值，将散列值作为受限伪装数据填充在观测值的零值位置，在汇聚节点处对接收到的隐私数据集剔除伪装数据后再次求取散列函数值，既增强了感知数据的机密性保护，又验证了完整性，同时也降低了数据融合传输过程中的通信开销。

2.1　感知数据的分布式压缩观测方法

分布式压缩感知（distributed compressed sensing,DCS）是一种利用感知数据之间的时空相关性，对感知数据进行高效压缩

观测、合理稀疏表示的数据融合方法。在分布式信源编码中，编码端无论是独立编码还是联合编码，经过传输后，在解码端都能进行联合解码，都可以得到相等的信息量[17]。在感知数据的分布式压缩观测阶段，首先将感知节点收集到的数据进行分布式压缩观测。

建立能够描述、处理感知数据的分布式压缩感知联合稀疏模型（JSM），在融合节点将附近通信范围内多个感知节点收集到的原始多媒体感知数据进行分布式压缩感知观测：

$$X_j = \Psi\theta_j = Z_c + Z_j \;,\quad j \in \{1, 2, \cdots, J\} \tag{1}$$

其中：J 为融合节点附近通信覆盖范围内的感知节点的个数；θ_j 为节点 j 感知数据的稀疏系数；Ψ 为 $N \times N$ 维公共稀疏基；Z_C 为感知数据 X_j 的公共稀疏成分；Z_j 分别为感知数据 X_j 的独立稀疏成分。

$$Z_C = \Psi\,\Theta_C \;,\quad \|\Theta_C\|_0 = K_C \tag{2}$$

$$Z_j = \Psi\,\Theta_j \;,\quad \|\Theta_j\|_0 = K_j \tag{3}$$

其中：式（2）（3）分别表示原始感知数据 X_j 的公共稀疏成分 Z_C、独立稀疏成分 Z_j 在稀疏基 Ψ 上是 K_C、K_j 稀疏的。

采用二进制随机观测矩阵对感知数据进行稀疏观测：

$$\begin{cases} y_1 = \Phi_1 x_1 = \Phi_1\Psi(\Theta_C + \Theta_1) \\ y_2 = \Phi_2 x_2 = \Phi_2\Psi(\Theta_C + \Theta_2) \\ \quad\vdots \quad \vdots \quad \vdots \quad \vdots \\ y_j = \Phi_j x_j = \Phi_j\Psi(\Theta_C + \Theta_j) \end{cases} \tag{4}$$

其中：y_j 为原始多媒体感知数据的 $M \times 1$ 维分布式压缩感知观测值向量；Φ_j 为 $M \times N$ 维稀疏二进制随机观测矩阵；x_j 为 j 节点的 $N \times 1$ 原始感知数据；Ψ 为 $N \times N$ 维公共稀疏基；Θ_C 为 j 节点的公共稀疏系数；Θ_j 为独立稀疏系数。

观测过程实际就是利用 $M \times N$ 维稀疏二进制随机观测矩阵 Φ_j 的 M 个行向量对稀疏系数向量进行投影，保留了信号重构所需的信息。原始多媒体感知数据观测值的位置集合用 *NTPS* 表示。

2.2 感知数据的机密性保护

单向散列函数，又称单向 hash 函数，是把任意长度的输入信息串转换为固定长度的输出串并且由输出串难以逆向求解得到输入信息串的一种函数。本文采用 MD5 单向散列函数，该函数把任意长度的输入信息串转换为 32 位的输出串，对分布式压缩感知观测数据利用 MD5 单向散列函数求散列值：

$$H_j = hash(y_j) \tag{5}$$

其中：H_j 为原始多媒体感知数据观测值的散列值，y_j 为输入的原始多媒体感知数据观测值。

原始多媒体感知数据的分布式压缩感知观测值由公共稀疏部分、独立稀疏部分、零值部分三部分组成，定义散列值为受限制的伪装数据，在观测值的零值部分填充散列值 H_j，并记录填充的位置，位置集合用 *RCPS* 表示。本文定义随机函数随机生成值域范围内的任意值作为非受限制的伪装数据，定义不受限制的伪装数据填充观测值零值的位置集合为 $\overline{RCPS}$。当观测值的零值数小于 32 时，受限的伪装数据在将零值位置填充满之后，剩下的受限伪装数据直接放在观测值的最后面；当观测值的零值数大于等于 32 时，可以将受限制的伪装数据全部填充到零值位置，并根据机密性保护需求的强度，进一步插入相应数量的非受限制的伪装数据。一般情况下，感知数据经过压缩感知后零值数较多，所以可以同时将受限制的伪装数据和非受限制的伪装数据填充到观测值的零值位置，增强对感知数据的机密性保护。

将受限制的伪装数据填充到观测值零值的位置信息 *RCPS* 和非受限制的伪装数据填充到观测值零值的位置信息 $\overline{RCPS}$ 共同描述为伪装数据填充到观测值的位置集合 *GCPS*，并将 *GCPS* 独立、加密地传给汇聚节点。同时，*GCPS* 和原始感知数据观测值的位置集合 *NTPS* 共同组成感知隐私数据包 I，即

$$GCPS = RCPS + \overline{RCPS} \tag{6}$$

$$(NTPS + RCPS + \overline{RCPS}) \subset I \tag{7}$$

本文将在一定通信范围内的感知节点化为一个簇，每个簇包含一个簇头节点即融合节点，由它负责对感知数据进行簇内融合，离汇聚节点更近的簇头节点可以充当中继节点对数据进一步完成簇间融合并传输，根据簇头节点距离汇聚节点的远近不同，融合后的数据可以经过一跳或者多跳传输到汇聚节点。在实践中被多次证明，对感知数据进行分布式压缩感知与直接对感知数据进行加密，在隐私保护中具有类似的效果[18]。感知数据观测及插入伪装数据的过程如图 1 所示。

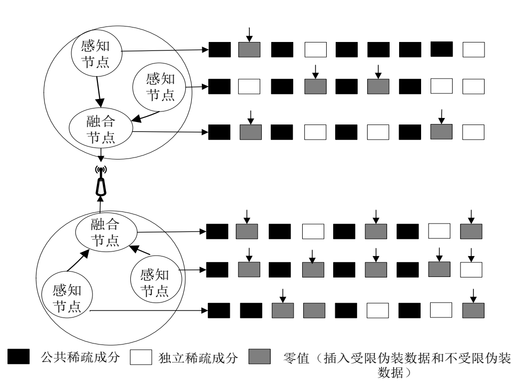


图 1　感知数据观测及插入伪装数据过程示意图

Fig.1　Process of perceptual data observation and inserting disguised data

2.3 感知数据的融合传输

一般情况下，感知数据融合计算可以根据实际应用需要选择加法、乘法、求平均值、最大值、最小值等进行综合操作。由于加法操作在隐私保护算法中限制因素较少，且其他几种操作也都可以转换为加法融合的形式，本文在感知数据融合传输过程中采用加法操作，加法操作如下所示：

$$\overline{Sum} = \overline{y_1} + \overline{y_2} + \cdots + \overline{y_j} \tag{8}$$

其中：$\overline{y_j}$ 为 j 节点对原始多媒体感知数据的观测值插入伪装数据之后的值。

2.4 感知融合数据的完整性验证

单向散列函数验证感知数据是否完整的过程如图 2 所示。

a)融合节点对感知数据进行压缩观测，利用单向散列函数对观测值求取散列值 HI；

b)汇聚节点在收到隐私数据集后，根据收到的插入伪装数据的位置信息剔除伪装数据，对观测值求取散列值 HO；

c)比较两个散列值，如果 $HO=HI$，说明感知数据在传输过程中是安全的，通过了完整性验证，下一步进行感知数据的重构；否则，说明数据在传输的过程中受到了攻击，目的节点收到的数据并不是真实的感知数据值，将之丢弃。

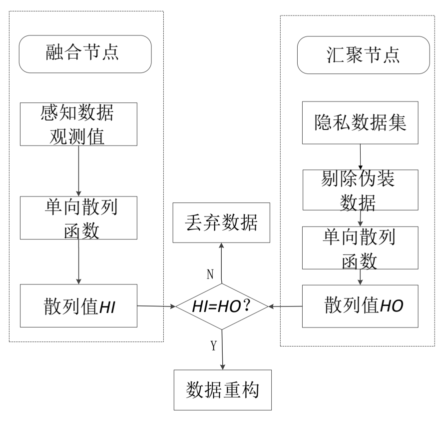


图 2　感知数据完整性验证过程

Fig.2　Integrity verification of perceptual data

2.5　感知融合数据的重构

对通过完整性验证的感知融合数据进行重构，恢复出原始感知数据。记分布式压缩感知信息算子为 $A^{DCS}=\Phi\Psi$，则 $y_j=\Phi\Psi\theta_j=A^{DCS}\theta_j$。文献[19]指出，如果 A^{DCS} 能够满足测量矩阵 Φ 与稀疏基矩阵 Ψ 不相关，那么 θ_j 可以通过求解最优 l_0 范数问题精确重构，重构如下所示：

$$\hat{\theta}_j=\arg\min\|\theta_j\|_0$$
$$s.t.\quad y_j=A^{DCS}\theta_j \tag{9}$$

其中：$\hat{\theta}_j$ 为汇聚节点收到的 j 节点的感知数据通过最优化方法求解得到的估计值；θ_j 为 j 节点感知数据的稀疏系数；y_j 为汇聚节点剔除感知数据分布式压缩观测值中的伪装数据，并经过完整性检验的数据。

最终恢复出的原始感知数据 $\hat{x}_j$ 为

$$\hat{x}_j=\Psi\hat{\theta}_j \tag{10}$$

2.6　基于分布式压缩感知和散列函数的数据融合隐私保护算法

在上述理论的基础上，给出基于分布式压缩感知和散列函数的数据融合隐私保护算法，以达到兼顾群智感知网络数据融合传输过程中对感知数据的机密性保护和完整性保护，同时减少数据融合传输网络通信量的目的。

基于分布式压缩感知和散列函数的数据融合隐私保护算法如下：

a)采用分布式压缩感知方法对簇内感知节点采集到的感知数据进行稀疏观测。

b)利用单向散列函数求取观测值的散列值 HI，并根据不同的隐私保护需求通过随机函数生成不受限的伪装数据，将散列值和不受限的伪装数据填充在观测数据的零值位置形成隐私数据包。将伪装数据的填充位置数据单独加密发给汇聚节点。

c)在中继节点处进一步完成簇间融合，形成隐私数据集。根据融合节点距离汇聚节点的远近不同，融合数据可经过一跳或多跳到达汇聚节点。

d)汇聚节点对剔除伪装数据的隐私数据集中的感知数据求取散列值 HO，比较两个散列值是否相等，如果 $HO=HI$，说明数据在传输的过程中是安全的，通过了完整性检验，对感知数据联合重构出原始数据；否则，说明数据在传输的过程中发生了改变，目的节点收到的数据并不是真实值，将之丢弃。

3　实验结果与分析

3.1　实验环境

本文算法采用 MATLAB 实现，仿真数据集采用标准的 Foreman、Hall Monitor 和 News 视频序列，三个视频序列的长度均为 300 帧，每帧图像的分辨率均为 256×256。

为了测试本文算法的性能，在相同条件下，将本文算法 DAP-DCSHF 与隐私数据融合完整性检测算法 ICKPDA 算法、基于分布式压缩感知的数据融合隐私保护算法 DCSPDA、隐私数据完整性保护算法 iPDA 进行仿真对比。分别从数据机密性保护水平、数据完整性保护水平、通信消耗三方面评估本文所提算法的有效性。DAP-DCSHF 算法在仿真中的参数设置如表 1 所示。

表 1 算法参数设置

Table 1　Algorithm parameter

参数	值	参数	值
网络面积（m²）	4500×3400	节点缓存（M）	30
节点通信方式	Bluetooth	Hash 值长度（位）	32
视频序列	Foreman，Hall Monitor，News	每个视频序列帧数（帧）	300
图像帧像素	256×256	链路破解概率	0.1-0.5
簇大小	3-15	簇数	30
关键帧采样率	0.7	非关键帧采样率	0.3

3.2　数据机密性保护性能分析

为了客观分析 DAP-DCSHF 算法对感知数据融合传输过程中的机密性保护效果，假设向真实观测数据中填充的散列值和不受限的伪装数据的位置信息都安全、保密的传输到了汇聚节点。记 DAP-DCSHF 真实感知数据集合的长度为 $|NTPS|$，节点的真实感知数据稀疏成分与受限伪装数据组成的节点秘密集合为 NSS，则 NSS 的补集的长度为 $|\overline{NSS}|$。假设每条链路被破解的概率为 q，该条链路上传输的感知数据被泄露的概率用 $P_{DAP-DCSHF}(q)$ 表示。泄露概率可以定义为当给定链路被破解时，

攻击者能从截获的隐私保护数据中成功恢复出真实感知数据的概率为

$$P_{DAP-DCSHF}=p\cdot\frac{1}{C_{|NTPS|+|\overline{NSS}|}^{|NTPS|}}=P\cdot\frac{|NTPS|!\cdot|\overline{NSS}|!}{(|NTPS|+|\overline{NSS}|)!} \quad (11)$$

攻击者只有获得簇内每个感知节点发送到簇头节点信道上的数据，才能获得某一个簇内的融合数据。假设簇内感知节点的个数为 n，则簇内融合数据隐私泄露的概率 $Prob_{DAP-DCSHF}$ 可以表示为

$$Prob_{DAP-DCSHF}=(P_{DAP-DCSHSF})^{n-1} \quad (12)$$

在 DCSPDA 算法中，攻击者如果想要获得簇内的融合数据，必须先获得每个感知节点发送到簇头节点的隐私数据包，才可能恢复出簇内的融合数据。所以，DCSPDA 算法簇内融合数据隐私泄露概率的计算与 DAP-DCSH 算法类似。

在 iPDA 算法中，每个感知节点首先将自己感知到多媒体数据分为 $2L$ 个切片，自己保留其中一片，并将剩下的 $2L-1$ 片数据分发给邻居节点，同时，它也会收到邻居节点发来的 n 个数据分片。攻击者如果想要破解感知数据，那么不仅需要得到该节点发送出去的 $2L-1$ 个数据分片，而且需要破解邻居节点发来的 n 个数据切片的链路，假设簇内感知节点数为 n，则数据遭到泄露概率的 $Prob_{iPDA}$ 可以表示为

$$Prob_{iPDA}=q^{2L-1}\cdot\sum_{k=0}^{n}P(RecJ=k)q^{k} \quad (13)$$

其中：$P(RecJ=k)$ 为该感知节点收到 k 个数据切片的概率。若 x 个节点收到 k 个切片，则

$$P(RecJ=k)=x/N \quad (14)$$

其中：N 为切片总数，由于 iPDA 算法需要构建两棵融合树，所以取 $L=2$，那么 $N=3$。

在 ICKPDA 算法中，攻击者只有把节点的两个私密种子和节点的度链接都破解，才能获得隐私数据。因此，隐私数据遭到泄露的概率 $Prob_{ICKPDA}$ 可以表示为：

$$Prob_{ICKPDA}=q^{2}\cdot\sum_{k=1}^{n}P(\deg=k)\cdot q^{k} \quad (15)$$

其中：n 为簇的大小，$P(\deg=k)$ 为度数为 k 的节点的概率。

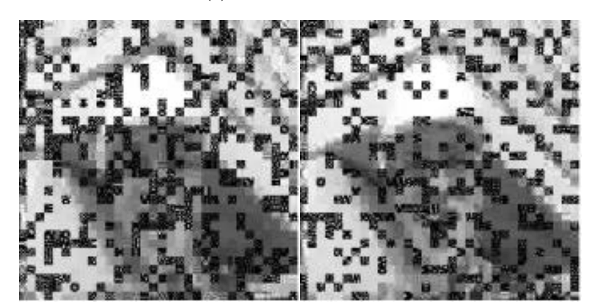
为了直观地体现不同算法在感知数据传输过程中的机密性保护性能，本文在簇大小 n 为 3 和链路泄露概率 q 为 0.1 且采用相同原图片帧的条件下，对不同算法造成数据泄露的后果进行仿真，其结果如图 3 所示。在采用 DAP-DCSHDF 算法和 DCSPDA 算法时，泄密的感知数据并不能完整地重构出原始的感知图像帧，而采用 ICKPDA 和 iPDA 算法时，泄露的感知数据有可能泄露原始感测图像的部分视觉特征。本文提出的 DAP-DCSHDF 算法相较于其他算法重构出的图像帧较为模糊，说明本文算法泄露的原始数据量较少，机密性保护效果较好。

DAP-DCSHF、DCSPDA、ICKPDA、iPDA 四种隐私保护数据融合算法在簇大小分别 3、5、8 的情况下，隐私数据泄露概率情况如图 4 所示，为了减小偶然因素带来的仿真结果误差，本文算法中的仿真值取 10 次仿真结果的平均值。

由图 4 可知，本文提出的算法和 DCSPDA 算法保护感知数据机密性的性能明显优于另外两种算法。因为 iPDA 和 ICKPDA 两种算法都是通过采用数据切片重组的方式实现数据融合机密性保护，机密性保护效果依赖于分片数的多少，如果增加分片数来增强数据融合的机密性保护，那么同时会增大通信开销。而 DAP-DCSHF 和 DCSPDA 算法都是基于分布式压缩感知对原始感知数据进行稀疏观测，分布式压缩感知的测量过程等效于隐私数据的加密过程，而且两种算法又同时在稀疏观测值的零值部分填充了伪装数据，增强了机密性保护。同时，本文提出的 DAP-DCSHF 的机密性保护性能略优于 DCSPDA 算法，因为 DCSPDA 中受限制的伪装数据的取值范围要比观测值小，不能有效的隐藏真实数据，而 DAP-DCSHF 中受限制的伪装数据为散列值，取值范围并没有受到观测值的影响。



(a)Foreman 原图像帧



(b)DAP-DCSHF　(n =3/ q =0.1)　　(c)DCSPDA　(n =3/ q =0.1)





(d)ICKPDA　(n =3/ q =0.1)　　(e)iPDA　(n =3/ q =0.1)

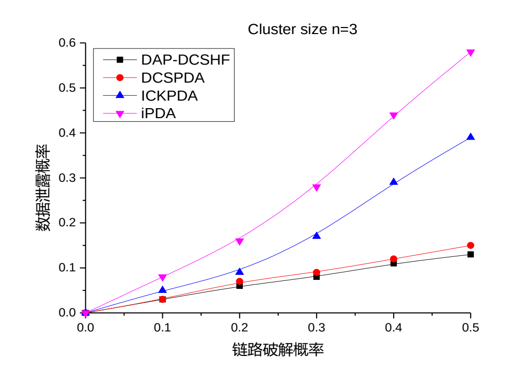
图 3　四种算法中泄密对感知数据的影响

Fig.3　EFFect of leaks on perceived data in four algorithms

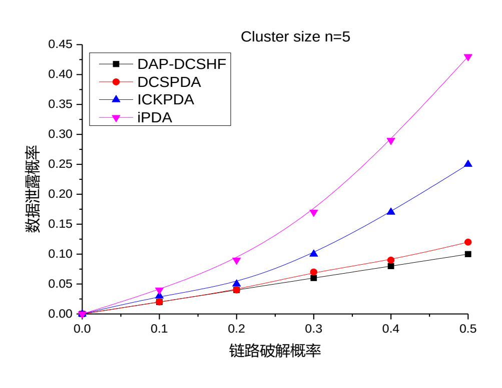
3.3　数据完整性保护性能分析

DCSPDA 算法增强了数据融合过程中的机密性保护，降低了通信开销，但是针对攻击者通过重放、伪造、窜改数据等威胁感知数据安全的行为并没有提出相应的解决方案。本文通过

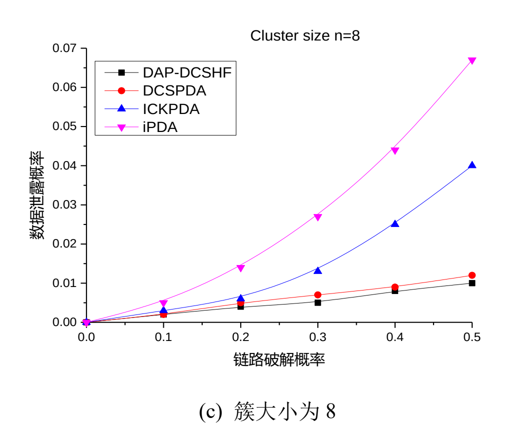
对 DAP-DCSHF、ICKPDA、iPDA 三种算法抵御重放、伪造、窜改数据攻击的程度分析算法保护数据完整性的性能。



(a)簇大小为 3



(b)簇大小为 5



(c) 簇大小为 8

图 4　簇大小 n=3, n=5, n=8 时，四种算法的数据泄露概率

Fig.4　Four algorithm of probability of data leakage in different cluster size

3.3.1 重放数据攻击

当融合数据遭到攻击者重放攻击时：iPDA 算法中所建立的两棵不相交的红融合树和蓝融合树，其中一颗树的融合值会改变，而另一颗树的融合值并没有发生改变，由于两颗树的融合结果不相等，所以 iPDA 算法可以很容易的检测出重放攻击。ICKPDA 算法中对最终得到的二元数据进行关联性计算后，得到的两个数据的还原结果并没有变，所以 ICKPDA 算法无法检测出重放攻击。DAP-DCSHF 算法中，因为重放攻击并没有使感知数据的观测值发生改变，所以不容易察觉是原数据还是重放数据。

3.3.2 伪造数据包攻击

当融合数据遭到攻击者伪造数据包攻击时：在 iPDA 算法中，如果攻击者只在其中一棵树中伪造了数据包，那么两颗树的融合结果不相等，算法很容易能够检测出来。但是如果攻击者对两棵融合树都伪造了数据包，由于两棵树的融合结果依然相等，iPDA 算法不能检测出此类攻击。

在 ICKPDA 算法中，汇聚节点得到的二元数据融合值也都相应增大，经过完整性检测值的计算后，两个值依然相等，所以，ICKPDA 算法不能抵抗伪造数据包攻击。DAP-DCSHF 算法中，对汇聚节点收到的隐私数据包剔除伪装数据后，汇聚节点收到的观测值所求取得散列值与原始数据观测值的散列值不相等，不能通过完整性验证，会被汇聚节点丢弃，所以，DAP-DCSHF 算法可以有效抵抗伪造数据包攻击。

3.3.3 窜改数据攻击

当融合数据遭到攻击者窜改数据攻击时：在 iPDA 算法中，与遭到伪造数据包攻击类似，如果只有其中一棵树遭到了窜改数据，那么算法很容易能够检测出来。但是如果两棵树都遭到了窜改数据，iPDA 算法便不能检测出。

在 ICKPDA 算法中，攻击者窜改了二元数据中的任何一元数据，都会直接导致二元数据之间的关联性遭到破坏，从而在汇聚节点处被检测出来，所以 ICKPDA 算法能够很好的抵御窜改数据的攻击。在 DAP-DCSHF 算法中，会造成观测值的散列值发生改变，所以 DAP-DCSHF 算法能够有效的检测出窜改数据的攻击。本文算法 DAP-DCSHF 与 DCSPDA、ICKPDA、iPDA 三种算法抵御重放、伪造、窜改数据攻击的性能分析（见表 2）。

表 2 抵御攻击性能分析

Table 2　Performance analysis of defense against attack

	重放数据攻击	伪造数据包攻击	窜改数据攻击
DAP-DCSHF	×	↑	↑
DCSPDA	×	×	×
ICKPDA	×	×	↑
iPDA	↑	×	↓

↑ 表示抵御攻击性能较强, ↓ 表示抵御攻击性能较弱, × 表示不能抵御攻击。

3.4　感知数据融合传输过程中的通信能耗

文献[20]表明感知节点的通信能耗远远大于计算能耗，所以，本节对感知数据融合传输过程中的通信能耗进行仿真分析。首先分析四种算法中每个节点要发送的数据包的数量：

在 DCSPDA 算法中，融合节点需要向汇聚节点发送一次伪装数据填充观测值零值部分的位置信息，并将隐私保护数据包传给上层中继节点，最终传输给汇聚节点。因此，DCSPDA 的通信开销为 $O(3n)$ 。由于 DCSPDA 算法与 DCSPDA 算法都是基于分布式压缩感知的隐私保护数据融合算法，所以通信开销大致相同。

iPDA 算法通过分片和重组数据来保护感知数据的机密性，为了保护数据的完整性，需要建立两棵不相交的红融合树和蓝融合树，每个节点需要分别向两棵融合树的邻居节点发送 $2l-1$ 片（l 为分片个数）数据，并且需要发送一个数据包用于建立融

合树，最后再将重组后的数据包发送出去。因此，iPDA 算法的数据通信开销为 $O((2l+1)n)$，为了保护数据的机密性，必须使 $l\geq 2$，因此，该算法的数据通信开销至少为 $O(5n)$。

在 ICKPDA 算法数据融合阶段，节点首先将数据分为 l 片，自己存储其中一片，其余的 $l-1$ 片数据加密后分别发送给不同的邻居节点，邻居节点收到数据切片后，先解密再与自身的数据进行融合；同时，该节点也会收到其他节点发来的数据分片，与自己的另一个数据分片进行融合。最后，所有节点都进行循环融合之后，上传融合数据到上层节点。由于分片数 l 必须大于等于 2，所以该算法的通信开销至少为 $O(4n)$。

为了合理比较四种算法的性能，将相同数量的感知数据传输到汇聚节点，比较四种算法的总数据通信量（数据包数），仿真结果如图 5 所示。

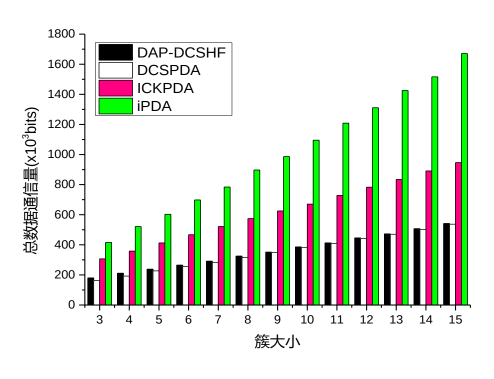


图 5　通信开销性能

Fig.5　Communication overhead

由图 5 可见，相比于 DAP-DCSHF 和 DCSPDA 算法，iPDA 和 ICKPDA 两种算法的通信开销明显较高，而且随着簇节点数的逐渐增加，算法之间的通信开销差别越来越大。因为 iPDA 算法为了对数据进行机密性保护，需要对同一个数据进行切片、重组，为了验证数据的完整性，建立了两颗不相交的红融合树和蓝融合树，并且对两棵树分别进行数据融合，导致冗余数据加倍增长。ICKPDA 算法虽然和 iPDA 算法一样采用数据分片重组的方式保护感知数据的机密性，但是由于同一个数据只需要融合一次，所以通信量比 iPDA 算法小很多。DAP-DCSHF 和 DCSPDA 算法利用感知数据之间的时空相关性，采用分布式压缩感知去除冗余数据，采集感知数据的重要稀疏成分，极大地减小了数据的通信开销，且随着簇内节点数的增多，这种优势越来越明显。由于 DAP-DCSHF 算法在观测值中填充的受限伪装数据为 32 位固定长度的散列值，而 DCSPDA 算法中是根据机密性保护需求强度动态调整受限伪装数据的长度，所以当簇节点数较小时，DAP-DCSHF 算法的通信量略高于 DCSPDA 算法，但是随着簇节点的增多，这两种算法的通信量趋近于一致。

4　结束语

针对群智感知网络数据融合传输过程中，感知数据隐私泄露严重，现有的算法以数据的机密性保护为主，而忽略了数据的完整性保护，同时能够兼顾数据机密性和完整性的隐私保护方案数据通信量又太高、保护强度较弱的问题，本文提出了一种基于分布式压缩感知和散列函数的数据融合隐私保护算法 DAP-DCSHF。该算法采用分布式压缩感知方法对感知数据进行稀疏观测，去除冗余数据，采用单向散列函数求取感知数据观测值的散列值，将散列值作为受限伪装数据和根据感知数据机密性保护需求的强度由随机函数生成的不受限制的伪装数据一起填充到感知数据观测值的零值部分，形成隐私数据包，增强感知数据的机密性保护。数据经过加性融合形成隐私数据集传输到汇聚节点，在汇聚节点对隐私数据集剔除伪装数据，再次求取散列值验证感知数据的完整性，对通过完整性验证的观测值通过重构恢复出原始感知数据。通过与隐私数据融合完整性保护保护算法 ICKPDA、基于分布式压缩感知的隐私保护数据融合算法 DCSPDA、兼顾完整性和机密性保护的算法 iPDA 对比，本文所提算法的机密性保护性能均优于三种算法；在防止攻击者伪造、窜改数据的完整性保护性能中均优于其他三种算法；在防止重放数据攻击中，保护性能略差于 iPDA 算法；在降低通信能耗方面，本文所提算法明显优于 ICKPDA、iPDA 算法，当数据量较小时，本文算法的通信量略高于 DCSPDA 算法，但是随着簇节点的增多，两种算法的通信量趋近于一致。本文算法兼顾了数据机密性和完整性保护，同时大大减少了网络中的通信开销，在实际应用中具有很强的适用性和可扩展性。当然，该算法也存在一些不足，例如本文仅适用于单一场景下感知数据融合隐私保护问题，但是在实际应用中，智能设备集成的传感器种类越来越多，考虑这种更复杂场景下的感知数据融合隐私保护问题，是未来工作的重点。

参考文献：

[1] 何宏，向朝参，肖书成，等．群智感知网络研究现状与发展 [J]. 吉林大学学报：信息科学版, 2016, 34 (3): 374-383. (He Hong, Xiang Chaocan, Xiao Shucheng, *et al.* Survey on crowd-sensing netwoks [J]. Journal of Jilin University: Information Science Edition, 2016, 34 (03): 374-383.)

[2] Sun W, Liu J. Congestion-aware communication paradigm for sustainable dense mobile crowd-sensing [J]. IEEE Communications Magazine, 2017, 55 (3): 62-67.

[3] Boubiche S, Boubiche D E, Bilami A, *et al.* Big data challenges and data aggregation strategies in wireless sensor networks [J]. IEEE Access, 2018, 2018 (6): 20558-20571.

[4] 曾菊儒，陈红，彭辉，等．参与式感知隐私保护技术 [J]. 计算机学报, 2016, 39 (3): 595-614. (Zeng Juru, Chen Hong, Peng Hui, *et al.* Privacy preservation in mobile participatory sensing [J]. Chinese Journal of Computers, 2016, 39 (3): 595-614.)

[5] Zhao Xiaomin, Zhu Jiabin, Liang Xueli, *et al.* Lightweight and integrity-protecting oriented data aggregation scheme for wireless sensor networks [J]. IET Information Security, 2017, 11 (2): 82-88.

[6] Ma Teng, Liu Yun, Zhang Zhenjiang. An energy-efficient reliable trust-based data aggregation protocol for wireless sensor networks [J].



International Journal of Control and Automation, 2015, 8 (1): 249-253.

[7] Akila V, Sheela T. Preserving data and key privacy in Data Aggregation for Wireless Sensor Networks [C]// Proc of the 2nd International Conference on Computing and Communications Technologies. Piscataway, NJ: IEEE Press, 2017: 282-287.

[8] Wu Dapeng, Yang Boran, Wang Ruyan. A scalable privacy-preserving big data aggregation method [J]. Digital Communications and Networks, 2016, 2 (3): 122-129.

[9] Lyu L J, Law Y W, Jin J, *et al.* Privacy-Preserving Aggregation of Smart Metering via Transformation and Encryption [C]// Proc of IEEE Trustcom//BigDataSE//ICESS. Piscataway, NJ: IEEE Press, 2017: 472-479.

[10] Wu Dapeng, Yang Boran, Wang Honggang, *et al.* Privacy-preserving multimedia big data aggregation in large-scale wireless sensor networks [J]. ACM Trans on Multimedia Computing Communications and Applications, 2016, 12 (4): 1-19.

[11] Zhang Jun, Zhu Jianghao, Jia Zongpu, *et al.* A secret confusion based energy-saving and privacy-preserving data aggregation algorithm [J]. Chinese Journal of Electronics, 2017, 26 (4): 740-746.

[12] Vinodha D, Mary Anita E A. Secure data aggregation techniques for wireless sensor networks: a review [J]. Archives of Computational Methods in Engineering, 2018, 2018 (8): 1-21.

[13] 宋成，张明月，彭维平，等．基于安全多方计算的车载网隐私保护机制 [J]．北京邮电大学学报, 2017, 40 (3): 67-71. (Song Cheng, Zhang Mingyue, Peng Weiping, *et al.* Privacy protection mechanism based on secure multi-party computation in VANET [J]. Journal of Beijing University of Posts & Telecommunications, 2017, 40 (3): 67-71.)

[14] 苘大鹏，王臣业，杨武，等．低能耗的无线传感器网络隐私数据融合方法 [J]．清华大学学报：自然科学版, 2017, 2017 (2): 213-219. (Man Dapeng, Wang Chenye, Yang Wu, *et al.* Energy-efficient cluster-based privacy data aggregation for wireless sensor networks [J]. Journal of Tsinghua University: Science and Technology, 2017, 2017 (2): 213-219.)

[15] He Wenbo, Nguyen H, Liu Xue, *et al.* iPDA: An integrity-protecting private data aggregation scheme for wireless sensor networks [C]// Proc of Military Communications Conference. Piscataway, NJ: IEEE Press, 2008: 1-7.

[16] 周强．无线传感器网络安全数据融合技术研究 [D]．南京：南京邮电大学, 2014. (Zhou Qiang. Research on secure data aggregation Technology of wireless sensor networks [D]. Nanjing: Nanjing University of Posts and Telecommunications, 2014.)

[17] Wang Wen, Zhu Jinkang, Zhang Sihai, *et al.* Tradeoff between efficiency and delay of distributed source coding for uplink transmissions in machine type communications [C]// Proc of the 9th International Conference on Wireless Communications and Signal Processing. Piscataway, NJ: IEEE Press, 2017: 1-6.

[18] Qian Jianhua, Zhang Xueying. Compressive data gathering based on even clustering for wireless sensor networks [J]. Journal of Computer Applications, 2018, 38 (6): 1691-1697.

[19] Hampton J, Doostan A. Basis adaptive sample efficient polynomial chaos (BASE-PC) [J/OL]. Journal of Computational Physics, 2018, 2018 (371) . (2017-07-02) [2018-07-20]. http://doi. org/10. 1016/j. jcp. 2018. 03. 035.

[20] Arbi I B, Derbel F, Strakosch F. Forecasting methods to reduce energy consumption in WSN [C]// Proc of the 12th IEEE International Instrumentation and Measurement Technology Conference. Piscataway, NJ: IEEE Press, 2017: 1-6.